REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
	Technical Report	-

4. TITLE AND SUBTITLE
Markov Kernels and the Conditional Extreme Value Model

5a. CONTRACT NUMBER
W911NF-10-1-0289

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

6. AUTHORS
Sidney Resnick, David Zeber

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Cornell University
Office of Sponsored Programs
Cornell University
Ithaca, NY 14850 -2820

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
57856-MA.36

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
The classical approach to extreme value modelling for multivariate data is to assume that the joint distribution belongs to a multivariate domain of attraction. In particular, this requires that each marginal distribution be individually attracted to a univariate extreme value distribution. The domain of attraction condition may be

15. SUBJECT TERMS
Markov dependence, multivariate regular variation, tail chain, heavy tail.

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
UU	UU	UU	UU		Sidney Resnick
					19b. TELEPHONE NUMBER
					607-255-1210

Report Title

Markov Kernels and the Conditional Extreme Value Model

ABSTRACT

The classical approach to extreme value modelling for multivariate data is to assume that the joint distribution belongs to a multivariate domain of attraction. In particular, this requires that each marginal distribution be individually attracted to a univariate extreme value distribution. The domain of attraction condition may be phrased conveniently in terms of regular variation of the joint distribution on an appropriate cone. A more flexible model for data realizations of a random vector was proposed by Heffernan and Tawn [45], under which not all the components are required to belong to an extremal domain of attraction. Such a model accomodates varying degrees of asymptotic dependence between pairs of components. Instead of starting from the joint distribution, assume the existence of an asymptotic approximation to the conditional distribution of the random vector given one of the components becomes extreme. Combined with the knowledge that the conditioning component belongs to a univariate domain of attraction, this leads to an approximation the probability of certain risk regions.

When originally proposed, the focus was on conditional distributions. This approach presents technical difficulties regarding the choice of version but makes sense when dealing with Markov kernels. We place this approach in the more general approach using vague convergence of measures and multivariate regular variation on cones.

MARKOV KERNELS AND THE CONDITIONAL EXTREME VALUE MODEL

SIDNEY I. RESNICK AND DAVID ZEBER

Abstract. The classical approach to extreme value modelling for multivariate data is to assume that the joint distribution belongs to a multivariate domain of attraction. In particular, this requires that each marginal distribution be individually attracted to a univariate extreme value distribution. The domain of attraction condition may be phrased conveniently in terms of regular variation of the joint distribution on an appropriate cone. A more flexible model for data realizations of a random vector was proposed by Heffernan and Tawn [45], under which not all the components are required to belong to an extremal domain of attraction. Such a model accomodates varying degrees of asymptotic dependence between pairs of components. Instead of starting from the joint distribution, assume the existence of an asymptotic approximation to the conditional distribution of the random vector given one of the components becomes extreme. Combined with the knowledge that the conditioning component belongs to a univariate domain of attraction, this leads to an approximation the probability of certain risk regions.

When originally proposed, the focus was on conditional distributions. This approach presents technical difficulties regarding the choice of version but makes sense when dealing with Markov kernels. We place this approach in the more general approach using vague convergence of measures and multivariate regular variation on cones.

1. Overview

The classical approach to extreme value modelling for multivariate data is to assume that the joint distribution belongs to a multivariate domain of attraction. In particular, this requires that each marginal distribution be individually attracted to a univariate extreme value distribution. The domain of attraction condition may be phrased conveniently in terms of regular variation of the joint distribution on an appropriate cone; see Das and Resnick [4, Proposition 4.1].

A more flexible model for data realizations of a random vector was proposed by Heffernan and Tawn [10], under which not all the components are required to belong to an extremal domain of attraction. Such a model accomodates varying degrees of asymptotic dependence between pairs of components. Instead of starting from the joint distribution, Heffernan and Tawn assumed the existence of an asymptotic approximation to the conditional distribution of the random vector given one of the components, as that component becomes extreme. Combined with the knowledge that the conditioning component belongs to a univariate domain of attraction, this leads to an approximation for the joint distribution, given that one component is extreme (e.g., exceeds some high threshold). However, the focus on conditional distributions presents some technical difficulties regarding the choice of version.

This approach was subsequently formalized as the Conditional Extreme Value Model (CEVM) by Heffernan and Resnick [9] and Das and Resnick [4, 5] in terms of regular variation of the joint distributions, but taking place on a smaller cone than the one employed in multivariate extreme value theory. This is related to the concept of hidden regular variation; see Resnick [14].

We return to the formulation of Heffernan and Tawn [10] in terms of conditional distributions, placing it in a more formal context by drawing upon the theory for transition kernels in a domain

Date: October 1, 2012.

S. I. Resnick and D. Zeber were partially supported by ARO Contract W911NF-10-1-0289 and NSA Grant H98230-11-1-0193 at Cornell University.

of attraction developed in [15]. In particular, we assume that the dependence between a pair of random variables (X, Y) is specified by a transition kernel K; this is appropriate, for example, in cases where one variable can be modeled as an explicit function of the other. In order to better fit in with the study of extremes of a random vector, we extend the kernel domain of attraction condition used in [15] beyond standardized regular variation to accomodate general linear normalization in both the initial state and the distribution of the next state. We examine conditions under which this extends to a CEVM, when combined with a marginal domain of attraction assumption, and we derive explicit formulas for the CEV limit measure in different cases. Also, through a number of revealing examples, we explore the properties of the normalization functions, and technicalities surrounding the choice of version of the conditional distribution and the limit distribution G.

2. Background

We begin by presenting some necessary background material. First, we review the basics of extended regular variation, which features prominently in the formulation of the CEVM, as well as some concepts of univariate extreme value theory. We then introduce the Conditional Extreme Value Model formally and discuss its basic properties.

2.1. Extended Regular Variation.

Regular variation and extended regular variation is important in the mathematical description of extreme and conditional extreme value theory. Standard references include [2, 6, 12, 13, 17]. The pair of functions $a : (0, \infty) \mapsto (0, \infty)$ and $f : (0, \infty) \mapsto \mathbb{R}$ are *extended regularly varying* (ERV) with parameters $\rho, k \in \mathbb{R}$ if as $t \to \infty$,

$$\frac{a(tx)}{a(t)} \longrightarrow x^\rho \qquad \text{and} \qquad \frac{f(tx) - f(t)}{a(t)} \longrightarrow \psi(x), \qquad x > 0, \tag{2.1}$$

[6, Appendix B.2], where

$$\psi(x) = \begin{cases} k\rho^{-1}(x^\rho - 1) & \rho \neq 0 \\ k \log x & \rho = 0 \end{cases}. \tag{2.2}$$

We will write this as $a, f \in \mathrm{ERV}_{\rho,k}$. Thus, $a \in \mathrm{RV}_\rho$, the regularly varying functions of index ρ. A useful identity is

$$\psi(x^{-1}) = -x^{-\rho}\psi(x). \tag{2.3}$$

Note that this differs slightly from the usual definition of extended regular variation, which assumes $k = 1$. If $\phi(x) := \lim_{t\to\infty}(f(tx) - f(t))/a(t)$ exists for $x > 0$, then a is necessarily regularly varying, and $\phi \equiv \psi$, the function given in (2.2). Also, the convergences in (2.1) are locally uniform, implying that

$$\frac{a(tx_t)}{a(t)} \longrightarrow x^\rho \qquad \text{and} \qquad \frac{f(tx_t) - f(t)}{a(t)} \longrightarrow \psi(x) \qquad \text{whenever } \ x_t \to x > 0.$$

Furthermore, if $k \neq 0$ we obtain the following properties depending on the value of ρ. Recall the sign function $\mathrm{sgn}(u) = u/|u|\,\mathbf{1}_{\{u\neq 0\}}$.

- If $\rho > 0$, then $f \cdot \mathrm{sgn}(k) \in \mathrm{RV}_\rho$, and $f(t)/a(t) \to k/\rho$.
- If $\rho < 0$, then $f(\infty) = \lim_{t\to\infty} f(t)$ exists finite, $(f(\infty) - f) \cdot \mathrm{sgn}(k) \in \mathrm{RV}_{-|\rho|}$, and $(f(\infty) - f(t))/a(t) \to k/|\rho|$.
- If $\rho = 0$, i.e., a is slowly varying, then $f \in \Pi(a)$ (see [6, Appendix B.2]). Suppose $k > 0$. Then $f(\infty) \leq \infty$ exists. If $f(\infty) = \infty$, then $f \in \mathrm{RV}_0$ and $f(t)/a(t) \to \infty$. If $f(\infty) < \infty$, then $f(\infty) - f \in \mathrm{RV}_0$, and $(f(\infty) - f(t))/a(t) \to \infty$. If $k < 0$, then $-f$ has these properties.

2.2. **Domains of Attraction.** For $\gamma \in \mathbb{R}$, define $\mathbb{E}_\gamma = \{x \in \mathbb{R} : 1 + \gamma x > 0\}$. Observe that

$$\mathbb{E}_\gamma = \begin{cases} (-\gamma^{-1}, \infty) & \gamma > 0 \\ (-\infty, \infty) & \gamma = 0 \\ (-\infty, |\gamma|^{-1}) & \gamma < 0 \end{cases}. \tag{2.4}$$

The distribution F of a random variable Y is in the *domain of attraction* of an extreme value distribution G_γ for some $\gamma \in \mathbb{R}$, written $F \in D(G_\gamma)$, if there exist functions $a(t) > 0$ and $b(t) \in \mathbb{R}$ such that

$$F^t\big(a(t)y + b(t)\big) \longrightarrow G_\gamma([-\infty, y])$$

weakly as $t \to \infty$, where $G_\gamma([-\infty, y]) = \exp\{-(1 + \gamma y)^{-1/\gamma}\}$ for $y \in \mathbb{E}_\gamma$ [6, 12]. This can be reformulated in terms of the tail of the distribution F as

$$t\,\mathsf{P}\big[Y - b(t)/a(t) > y\big] \longrightarrow (1 + \gamma y)^{-1/\gamma}, \qquad y \in \mathbb{E}_\gamma. \tag{2.5}$$

If $\gamma = 0$, we interpret the limit as e^{-y}.

If (2.5) holds for some functions a and b, then it holds for [6, Theorem 1.1.6, p. 10])

$$b(t) = \left(\frac{1}{1 - F_Y}\right)^{\leftarrow}(t) = F_Y^{\leftarrow}\big(1 - t^{-1}\big), \tag{2.6}$$

where $g^{\leftarrow}$ denotes the left-continuous inverse of a nondecreasing function g. Hence, by inversion, (2.5) implies that

$$\frac{b(tx) - b(t)}{a(t)} \to \frac{x^\gamma - 1}{\gamma}\,\mathbf{1}_{\{\gamma \neq 0\}} + \log x\,\mathbf{1}_{\{\gamma = 0\}}, \tag{2.7}$$

i.e., $a, b \in \mathrm{ERV}_{\gamma,1}$. Furthermore, if functions $\tilde{a} > 0$ and $\tilde{b} \in \mathbb{R}$ on $(0, \infty)$ are *asymptotically equivalent* to a, b, i.e., they satisfy

$$\frac{\tilde{a}(t)}{a(t)} \longrightarrow 1 \qquad \text{and} \qquad \frac{\tilde{b}(t) - b(t)}{a(t)} \longrightarrow 0 \qquad \text{as } t \to \infty,$$

then (2.5) and (2.7) hold with a, b replaced by $\tilde{a}, \tilde{b}$. It follows that (2.5) is equivalent to $t\,\mathsf{P}[b^{\leftarrow}(Y) > ty] \to y^{-1}$ for $y > 0$, i.e., $1 - F_{b^{\leftarrow}(Y)} \in \mathrm{RV}_{-1}$. This is known as standardization (see [12, Chapter 5]). We say that Y is in the *standardized domain of attraction*, and write $F \in D(G_1^*)$, if

$$t\,\mathsf{P}[Y > ty] \longrightarrow y^{-1}, \qquad y > 0.$$

2.3. **The Conditional Extreme Value (CEV) Model.** Denote by $\overline{\mathbb{E}}_\gamma$ the closure on the right of the interval $\mathbb{E}_\gamma$. A bivariate random vector (X, Y) on $\mathbb{R}^2$ follows a *Conditional Extreme Value Model* (CEVM) if there exist a non-null Radon measure μ on $[-\infty, \infty] \times \overline{\mathbb{E}}_\gamma$, and functions $a(t), \alpha(t) > 0$, $b(t), \beta(t) \in \mathbb{R}$, such that, as $t \to \infty$,

$$t\,\mathsf{P}\left[\left(\frac{X - \beta(t)}{\alpha(t)}, \frac{Y - b(t)}{a(t)}\right) \in \cdot\,\right] \xrightarrow{v} \mu(\cdot) \qquad \text{in } \mathbb{M}_+([-\infty, \infty] \times \overline{\mathbb{E}}_\gamma), \tag{2.8}$$

where μ satisfies the *conditional non-degeneracy* conditions: for each $y \in \mathbb{E}_\gamma$,

$$\begin{aligned} &\mu([-\infty, x] \times (y, \infty]) \text{ is not a degenerate distribution in } x; \\ &\mu(\{\infty\} \times (y, \infty]) = 0. \end{aligned} \tag{2.9}$$

It is convenient to choose the normalization such that

$$H(x) := \mu([-\infty, x] \times (0, \infty]) \text{ is a probability distribution on } [-\infty, \infty]. \tag{2.10}$$

See [4, 9] for details and [10] for background.

Some remarks: By applying the joint convergence (2.8) to rectangles $[-\infty, \infty] \times (y, \infty]$, we see that the distribution of Y is necessarily attracted to G_γ for some γ. Also, an important property is that the functions α, β are ERV for some $\rho, k \in \mathbb{R}$ [9, Proposition 1]. The limit measure μ in (2.8) is a product measure if and only if $(\rho, k) = (0, 0)$ [9, Proposition 2].

Condition (2.9) is somewhat different from what is given in [4, 9] which contained a redundancy and allowed mass on the line $\{\infty\} \times (-\infty, \infty]$ through infinity. Mass on this line invalidates the convergence to types theorem [7, 11]. The theory in [4, 9] employs convergence of types arguments which require no mass on the lines through $\{\infty\}$. Condition (2.8) entails $Y \in D(G_\gamma)$ and $\mu([-\infty, x] \times \{\infty\}) = 0$. We require also that $\mu(\{\infty\} \times (y, \infty]) = 0$, a fact not implied by (2.8). Example 3.6 presents a case where (2.8) holds for two distinct normalizations which are not asymptotically equivalent, yielding a different limit measures. One of the limit measures has $\mu(\{\infty\} \times (y, \infty]) > 0$ and the other has $\mu(\{\infty\} \times (y, \infty]) = 0$.

3. Standard Case

Let (X, Y) be a random vector on $\mathbb{R}^2$, with dependence specified by a transition kernel K:

$$\mathsf{P}[X \in \cdot \,\big|\, Y = y] = K(y\,,\,\cdot) \qquad y \in \mathbb{R}.$$

We show if the distribution of Y is in an extremal domain of attraction, and K belongs to the domain of attraction of a probability distribution G (a notion to be defined precisely), then (X, Y) follows a CEVM.

We begin with the *standard case* which means that $(X, Y) \in [0, \infty)^2$, and $F_Y \in D(G_1^*)$,

$$tF_Y(t\,\cdot) \xrightarrow{v} \nu_1(\cdot) \quad \text{in } \mathbb{M}_+(0, \infty] \quad \text{as } t \to \infty, \tag{3.1}$$

and $K \in D(G)$ meaning

$$K(t\,,\,t\cdot) \Rightarrow G(\cdot) \quad \text{on } [0, \infty]. \tag{3.2}$$

Here $\Rightarrow$ denotes weak convergence, and we write $\xi \sim G$ to mean $G(\cdot) = \mathsf{P}[\xi \in \cdot]$.

3.1. Standard CEVM Properties.

Conditions(3.1) and (3.2) imply (X, Y) follows a CEVM, provided $G \neq \epsilon_0$, i.e., unit mass at $\{0\}$.

Theorem 3.1. *Suppose that the joint distribution of the random vector (X, Y) on $[0, \infty)^2$ satisfies (3.1) and (3.2), where G is a probability distribution on $[0, \infty)$. Then*

$$t\,\mathsf{P}\left[(X, Y) \in t\,\cdot\,\right] \xrightarrow{v} \mu(\cdot) \quad \text{in } \mathbb{M}_+([0, \infty] \times (0, \infty]), \tag{3.3}$$

with limit measure μ given by

$$\mu([0, x] \times (y, \infty]) = \int_{(y,\infty]} \nu_1(du)\,\mathsf{P}[\xi \le xu^{-1}] = \int_y^\infty G(x/u)\nu_1(du), \qquad x, y > 0. \tag{3.4}$$

Furthermore, μ satisfies the conditional non-degeneracy conditions (2.9) provided $G \neq \epsilon_0$.

Proof. The convergence (3.3) is an application of Proposition **??** (a) (p. **??**), with $\alpha = 1$ and $m = 1$, and with Y playing the role of X_0. From (3.5) below, we see that $\mu\left([0, x] \times (y, \infty]\right)$ is continuous in x, and not constant provided $G \neq \epsilon_0$. Also, since $\mu((x, \infty] \times (y, \infty]) = \int_{(y,\infty]} \nu_1(du)\,\mathsf{P}[\xi > xu^{-1}]$, that $\mu(\{\infty\} \times (y, \infty]) = 0$ follows from the fact that $G(\{\infty\}) = 0$. Therefore, μ satisfies (2.9). ☐

3.1.1. *Properties of the limit measure μ.*

By changing variables, μ can be expressed as

$$\begin{aligned} \mu([0, x] \times (y, \infty]) &= \frac{1}{x}\int_0^{x/y} \mathsf{P}[\xi \le u]du \\ &= y^{-1}\,\mathsf{P}[\xi \le x/y] - x^{-1}\,\mathsf{E}\,\xi\,\mathbf{1}_{\{\xi \le x/y\}}, \end{aligned} \tag{3.5}$$

showing that μ is continuous in x and y and if G has a density, then so does μ. Note that the continuity in (3.5) holds even if G is degenerate, i.e., $G = \epsilon_c$ for some $c > 0$; see Example 3.4 (p. 6).

Non-degeneracy of G only becomes relevant in the non-standard case. Moreover, μ cannot be a product measure [4, Lemma 3.1].

From (3.5) we also observe that the y-axis through the origin is assigned mass proportional to $G(\{0\})$ since $\mu(\{0\} \times (y, \infty]) = y^{-1}G(\{0\})$. Mass on vertical slices of space depends on $\mathsf{E}\,\xi$, since $\mu((x, \infty] \times (0, \infty]) = x^{-1}\,\mathsf{E}\,\xi \leq \infty$. In terms of conditional distributions, (3.3) implies

$$\mathsf{P}[X \leq tx \,\big|\, Y > t] \Rightarrow H(x) := \mu([0, x] \times (1, \infty]) = \frac{1}{x}\int_0^x \mathsf{P}[\xi \leq u]du.$$

3.1.2. *Extending to a larger cone.* Convergence (3.3) extends to standard regular variation on the larger cone $[0, \infty]^2 \setminus \{\mathbf{0}\}$, so that the distribution of (X, Y) is in a bivariate domain of attraction, if and only if $F_X \in D(G_1^*)$ as well [4, Proposition 4.1]. In this case,

$$t\,\mathsf{P}\left[t^{-1}(X, Y) \in [\mathbf{0}, (x, y)]^{\mathrm{c}}\right] \longrightarrow \frac{1}{x}\left(1 + \int_0^{x/y} \mathsf{P}[\xi \leq u]du\right), \tag{3.6}$$

implying that $\mathsf{E}\,\xi \leq 1$, and the x-axis receives mass according to $\mu((x, \infty] \times \{0\}) = x^{-1}(1 - \mathsf{E}\,\xi)$.

3.1.3. *Degenerate G; asymptotic independence.* If $G = \epsilon_0$, then the convergence (3.3) holds with limit measure $\mu([0, x] \times (y, \infty]) = y^{-1}$ but conditional non-degeneracy (2.9) fails, since all the mass lies on the y-axis, so (X, Y) does not follow a standard CEVM. This is in fact a manifestation of asymptotic independence. Indeed,

$$\mathsf{P}[X > tx \,\big|\, Y > t] \to 0$$

for any x, so, given that Y is extreme (exceeding the threshold $u(t) = t$), it is very unlikely to observe X to be similarly extreme. If the joint distribution of (X, Y) is regularly varying on the larger cone $[0, \infty]^2 \setminus \{\mathbf{0}\}$, then

$$t\,\mathsf{P}\left[t^{-1}(X, Y) \in [\mathbf{0}, (x, y)]^{\mathrm{c}}\right] \longrightarrow x^{-1} + y^{-1},$$

which means that X and Y are asymptotically independent in the usual sense [9, Section 5]. In this case, (X, Y) do not follow a standard CEVM because of degeneracy, although a CEVM may hold if X is normalized differently, as in Section 4.

This suggests viewing the parameter $G(\{0\})$ as quantifying the "degree" of asymptotic dependence from Y to X. For example, given Y, we could write X as a mixture

$$X = WX_0 + (1 - W)X_1, \tag{3.7}$$

where X_0 and Y are asymptotically independent, X_1 and Y are asymptotically dependent, and $W \sim \text{Bernoulli}(G(\{0\}))$. This relates the canonical form of the update function representation of K (see Section **??**, p. **??**). Asymptotic dependence in the reverse direction, given large X, would then be quantified by $1 - \mathsf{E}\,\xi$ if appropriate. The latter phenomenon is hinted at by Segers [16] in his definition of the "back-and-forth tail chain" to approximate stationary Markov chains .

3.2. **Examples.** Examples illuminate properties of the CEVM based on Markov kernels as in (3.2). First, given any distribution G on $[0, \infty)$, we can construct a CEVM whose limit measure μ is built on G as in (3.5). See [4, Example 8].

Example 3.1. Take G to be any probability distribution on $[0, \infty)$ not concentrating at 0. Let $Y \sim \text{Pareto}(1)$ on $[1, \infty)$, $\xi \sim G$, independent of Y, and put $X = \xi Y$. A version of the conditional distribution is

$$K(y\,,\,\cdot) = \mathsf{P}[X \in \cdot \,\big|\, Y = y] = \mathsf{P}[\xi Y \in \cdot \,\big|\, Y = y] = G(y^{-1}\cdot),$$

and K satisfies (3.2) and in fact and $K(t\,,\,t\cdot) = G(\cdot)$. Consequently, (X,Y) follows a standard CEVM with limit measure as in (3.4). In fact, for $x,y>0$, we have

$$\begin{aligned}\mathsf{P}[X \le x, Y > y] &= \int_{(y,\infty]} K(u\,,\,[0,x])P[Y \in du] \\ &= \int_{y\vee 1}^{\infty} \mathsf{P}[\xi \le xu^{-1}]u^{-2}du = \frac{1}{x}\int_0^{x\wedge\frac{x}{y}} \mathsf{P}[\xi \le u]du.\end{aligned}$$

Furthermore, (X,Y) belong to a standard bivariate domain of attraction (3.6) iff $F_X \in D(G_1^*)$ as well. The marginal distribution of $X = \xi Y$ is

$$F_X([0,x]) = \frac{1}{x}\int_0^x \mathsf{P}[\xi \le u]du = \frac{1}{x}\int_0^x G(u)du = H(x),$$

which has density $f_X(x) = x^{-1}\big\{G([0,x]) - H([0,x])\big\}$ for $x \ge 0$. Since

$$\lim_{t\to\infty} t\,\mathsf{P}[X > tx] = \lim_{t\to\infty}\frac{1}{x}\int_0^{tx}\mathsf{P}[\xi > u]du = x^{-1}\,\mathsf{E}\,\xi \quad (\le \infty),$$

(X,Y) belongs to the standard domain of attraction iff $\mathsf{E}\,\xi = 1$. □

Using the Example 3.1 recipe, we explore the CEVM in a variety of special cases.

Example 3.2. Choose $\xi \sim \text{Exp}(\lambda)$ and we have $X = \lambda^{-1}YE$, where $E \sim \text{Exp}(1)$. The limit measure is

$$\mu([0,x]\times(y,\infty]) = \frac{1}{x}\int_0^{x/y}(1-e^{-\lambda u})du = \frac{1}{y} - \frac{1}{\lambda x} + \frac{e^{-\lambda x/y}}{\lambda x},$$

and the marginal distribution of X is $F_X(x) = 1-(\lambda x)^{-1}(1-e^{-\lambda x})$ with density $f(x) = \lambda^{-1}x^{-2}(1-e^{-\lambda x}) - x^{-1}e^{-\lambda x}$, and F_X satisfies (3.1) iff $\lambda = 1$. □

Next, we suppose ξ is heavy-tailed.

Example 3.3. For $\alpha > 0$ let $\xi \sim \text{Pareto}(\alpha)$ so $\bar{G}(x) = x^{-\alpha}$, $x \ge 1$. The limit measure assigns no mass to $\{(x,y) : 0 \le x \le y\}$, and for $x > y > 0$,

$$\mu([0,x]\times(y,\infty]) = \begin{cases} \dfrac{1}{y} - \left(\dfrac{\alpha}{\alpha-1}\right)\dfrac{1}{x} + \dfrac{y^{\alpha-1}}{x^{\alpha}(\alpha-1)} & \alpha > 1 \\[2ex] \dfrac{1}{y} - \dfrac{1}{x} - \dfrac{\log x}{x} + \dfrac{\log y}{x} & \alpha = 1 \\[2ex] \dfrac{1}{y} + \left(\dfrac{2-\alpha}{1-\alpha}\right)\dfrac{1}{x} + \dfrac{1}{x^{\alpha}y^{1-\alpha}(1-\alpha)} & \alpha < 1. \end{cases}$$

When $\alpha \le 1$, $\mathsf{E}\,\xi = \infty$ and $\mu((x,\infty]\times(y,\infty]) = y^{-1} - \mu([0,x]\times(y,\infty]) \to \infty$ as $y \downarrow 0$. □

It is also possible that G is discrete, although the CEVM limit measure μ remains continuous.

Example 3.4. Suppose ξ has discrete distribution $\mathsf{P}[\xi = k] = a_k$, $k = 0,1,\ldots$. In this case, the limit measure is given by

$$\mu([0,x]\times(y,\infty]) = \frac{1}{x}\int_0^{x/y}\Big(\sum_{k=0}^{[u]} a_k\Big)du = \sum_{k=0}^{[x/y]} a_k(y^{-1} - kx^{-1}),$$

which is continuous in x and y, and $F_X(x) = \sum_{k=0}^{[x]} a_k(1-kx^{-1})$. In particular, if $\mathsf{P}[\xi = c] = 1$ for some $c > 0$, we obtain

$$\mu([0,x]\times(y,\infty]) = (y^{-1} - cx^{-1})\,\mathbf{1}_{\{x > cy > 0\}}\,.$$

The conditional non-degeneracy conditions (2.9) are satisfied even though G is degenerate. □

The final example shows how G reflects asymptotic independence between X and Y.

Example 3.5. Consider $Y \sim$ Pareto(1), and Z independent of Y such that $\mathsf{P}[Z < \infty] = 1$. Take

$$X = Y \vee Z = Y\,\mathbf{1}_{\{Y \geq Z\}} + Z\,\mathbf{1}_{\{Z > Y\}}\,.$$

Given that Y is extreme, it is very unlikely that Z is as extreme as Y, since they are independent. We have

$$K(y\,,\,[0,x]) = \mathsf{P}[Y \leq x, Z \leq x \,|\, Y = y] = \mathsf{P}[Z \leq x]\,\mathbf{1}_{\{x \geq y\}},$$

and so

$$K(t\,,\,t[0,x]) = \mathsf{P}[Z \leq tx]\,\mathbf{1}_{\{x \geq 1\}} \longrightarrow \epsilon_1([0,x]) = \mathbf{1}_{\{x \geq 1\}} = G([0,x]).$$

As in the previous example, the limit measure is

$$\mu([0,x] \times (y,\infty]) = (y^{-1} - x^{-1})\,\mathbf{1}_{\{x > y > 0\}}\,.$$

On the other hand, consider $X' = Y \wedge Z = Y\,\mathbf{1}_{\{Y < Z\}} + Z\,\mathbf{1}_{\{Z \leq Y\}}$. When Y is large, it is likely $X' = Z$, so X' is asymptotically independent of Y. Indeed, in this case,

$$K(y\,,\,(x,\infty]) = \mathsf{P}[Y > x, Z > x \,|\, Y = y] = \mathsf{P}[Z > x]\,\mathbf{1}_{\{y > x\}},$$

from which

$$K(t\,,\,t(x,\infty]) = \mathsf{P}[Z > tx]\,\mathbf{1}_{\{x < 1\}} \longrightarrow 0$$

for $x > 0$. Therefore, $G = \epsilon_0$, and the conditional non-degeneracy conditions do not hold. □

3.3. **Counter-examples.** As expected, the converse to Theorem 3.1 can fail. If (X, Y) follows a non-degenerate CEVM as in (3.3), and K is a specific version of the conditional distribution $\mathsf{P}[X \in \cdot \,|\, Y = y]$, it does not necessarily follow that there exists a distribution G such that (3.2) holds. The failure of (3.2) can happen in two ways. There may exist a probability distribution G on $[0,\infty]$ satisfying (3.2) with $G(\{\infty\}) > 0$ or it may be possible to obtain two distinct limit distributions down different subsequences $\{t_n\}$ and $\{t'_n\}$.

Example 3.6 where $G(\{\infty\}) > 0$ emphasizes the importance of assuming $\mu(\{\infty\} \times (y,\infty]) = 0$.

Example 3.6. As usual, take $Y \sim$ Pareto(1) and suppose that

$$X = WY + (1 - W)Y^2,$$

where $W \sim$ Bernoulli(p) independent of Y. Then

$$K(y\,,\,\cdot) = \mathsf{P}[X \in \cdot \,|\, Y = y] = p\epsilon_y + (1-p)\epsilon_{y^2},$$

so

$$K(t\,,\,t\,\cdot) = p\epsilon_1 + (1-p)\epsilon_t \Rightarrow p\epsilon_1 + (1-p)\epsilon_\infty = G \qquad \text{on } [0,\infty].$$

Indeed, for $0 \leq x < \infty$,

$$K(t\,,\,t[0,x]) = p\epsilon_1([0,x]) + (1-p)\epsilon_t([0,x]) \longrightarrow p\epsilon_1([0,x])$$

showing that $G(\{\infty\}) = 1 - p$.

On the other hand, for $x, y > 0$,

$$\begin{aligned} \mathsf{P}[X \leq x, Y > y] &= p\,\mathsf{P}[Y \leq x,\, Y > y] + (1-p)\,\mathsf{P}[Y^2 \leq x,\, Y > y] \\ &= p\left[\frac{1}{(y \vee 1)} - \frac{1}{x}\right]\mathbf{1}_{\{x \geq (y \vee 1)\}} + (1-p)\left[\frac{1}{(y \vee 1)} - \frac{1}{\sqrt{x}}\right]\mathbf{1}_{\{x \geq (y \vee 1)^2\}}, \end{aligned}$$

so for t sufficiently large,

$$\begin{aligned} t\,\mathsf{P}[X \leq tx, Y > ty] &= p\left[\frac{1}{y} - \frac{1}{x}\right]\mathbf{1}_{\{x \geq y\}} + (1-p)\left[\frac{1}{y} - \frac{\sqrt{t}}{\sqrt{x}}\right]\mathbf{1}_{\{\sqrt{x}/\sqrt{t} \geq y\}} \\ &\longrightarrow p\left[\frac{1}{y} - \frac{1}{x}\right]\mathbf{1}_{\{x \geq y\}} = \mu([0,x] \times (y,\infty]). \end{aligned} \tag{3.8}$$

The measure μ assigns positive mass to $\{\infty\} \times (y, \infty]$ since

$$\mu((x,\infty] \times (y,\infty]) = y^{-1} - \mu([0,x] \times (y,\infty]) = \frac{1}{y}\,\mathbf{1}_{\{x<y\}} + \left[\frac{1-p}{y} + \frac{p}{x}\right]\mathbf{1}_{\{x\geq y\}},$$

and thus $\mu(\{\infty\} \times (y,\infty]) = (1-p)y^{-1}$. Therefore, μ does not satisfy (2.9).

Under a different normalization, we obtain a proper limit G. Indeed, note that

$$K(t\,,\,t^2\cdot) = p\epsilon_{t^{-1}} + (1-p)\epsilon_1 \Rightarrow p\epsilon_0 + (1-p)\epsilon_1 \sim \text{Bernoulli}(1-p),$$

and hence,

$$\begin{aligned} t\,\mathsf{P}[X \leq t^2 x, Y > t\cdot y] =& p(y^{-1} - (tx)^{-1})\,\mathbf{1}_{\{x\geq y/t\}} + (1-p)(y^{-1} - x^{-1/2})\,\mathbf{1}_{\{x\geq y\}} \\ &\longrightarrow py^{-1} + (1-p)(y^{-1} - x^{-1/2})\,\mathbf{1}_{\{x\geq y\}}\,. \end{aligned}$$

This limit does satisfy (2.9). □

Without condition (2.9), the convergence of types theorem fails and it is possible to obtain different CEV limits under different normalizations. From (3.4), $\mu(\{\infty\} \times (y,\infty]) = G(\{\infty\})y^{-1}$ and excluding defective distributions in Theorem 3.1 avoids cases like the previous one.

Here is an example of a CEVM where the normalized kernel K does not have a unique limit.

Example 3.7. Suppose $Y \sim \text{Pareto}(1)$, and define X by

$$X = WY + (1-W)2Y\,\mathbf{1}_{\{Y\in[0,\infty)\setminus\mathbb{N}\}}$$

where $W \sim \text{Bernoulli}(p)$ independent of Y. In other words, given $Y = y$, X takes the value y or $2y$ according to a coin flip, unless y is an integer, in which case X will be either y or 0. The CEVM holds for (X, Y). Since $\mathsf{P}[Y \in \mathbb{N}] = 0$, we have

$$\begin{aligned} \mathsf{P}[X \leq x, Y > y] &= \mathsf{P}[X \leq x,\, Y > y,\, Y \in [0,\infty)\setminus\mathbb{N}] \\ &= p\,\mathsf{P}(Y \leq x, Y > y) + (1-p)\,\mathsf{P}(2Y \leq x, Y > y) \\ &= p(y^{-1} - x^{-1})\,\mathbf{1}_{\{x\geq y\}} + (1-p)(y^{-1} - 2x^{-1})\,\mathbf{1}_{\{x\geq 2y\}}, \end{aligned}$$

and $t\,\mathsf{P}[X \leq tx, Y > ty] = \mathsf{P}[X \leq x, Y > y]$, which satisfies (2.10) and the requirement that $\mu((\cdot) \times (y,\infty])$ not be degenerate for any y. However, the conditional distribution of X given Y is

$$K(y\,,\,\cdot) = \begin{cases} p\epsilon_y + (1-p)\epsilon_0 & y \in \mathbb{N} \\ p\epsilon_y + (1-p)\epsilon_{2y} & y \in [0,\infty)\setminus\mathbb{N} \end{cases},$$

so

$$K(t\,,\,t\,\cdot) = \begin{cases} p\epsilon_1 + (1-p)\epsilon_0 & t \in \mathbb{N} \\ p\epsilon_1 + (1-p)\epsilon_2 & t \in [0,\infty)\setminus\mathbb{N} \end{cases}.$$

We obtain different limits along the sequences $t_n = n$ and $t'_n = n/2$ and $K(t, t\,\cdot)$ does not converge. □

The technical difficulty highlighted in Example 3.7 is that conditional distributions of the form $\mathsf{P}[X \in \cdot \,|\, Y = y]$ are only specified up to sets of $\mathsf{P}[Y \in \cdot\,]$-measure zero. If Y is absolutely continuous, we can alter the conditional probability for a countable number of y without affecting the joint distribution. Consequently, it is difficult to construct a convergence theory based on conditional distributions. The best one can do is fix a version of the kernel or, if circumstances allow, choose a version of the kernel with some claim to naturalness based on smoothness. This is the reason the approach in [4, 9] is based on vague convergence of measures rather than convergence of conditional distributions as in [10].

4. General Normalization for X

The CEVM allows different normalizations for X and Y, as in (2.8), but the formulation $K \in D(G)$ in (3.2), imposes the same normalization for both. We now allow general linear normalizations of X in the kernel condition, continuing to assume $F_Y \in D(G_1^*)$ (3.1).

We will assume the following generalization of (3.2): there exists a non-degenerate probability distribution G on $[-\infty,\infty)$, scaling and centering functions $\alpha(t)>0$, $\beta(t)\in\mathbb{R}$, such that

$$K\big(t\,,\,[-\infty,\alpha(t)x+\beta(t)]\big) \Rightarrow G([-\infty,x]) \quad \text{on } [-\infty,\infty]. \tag{4.1}$$

4.1. **CEVM Properties.**

Consider the decomposition

$$t\,\mathsf{P}\left[\frac{X-\beta(t)}{\alpha(t)} \le x, Y > ty\right] = \int_{(y,\infty]} t\,\mathsf{P}[Y\in tdu]K\big(tu\,,\,[-\infty,\alpha(t)x+\beta(t)]\big). \tag{4.2}$$

By a variant of the continuous mapping theorem ([15, Lemma 8.2], this will converge provided $K\big(tu(t)\,,\,[-\infty,\alpha(t)x+\beta(t)]\big)\to\varphi_x(u)$ whenever $u(t)\to u>0$. What are conditions such that kernel convergence (4.1) implies joint distribution convergence in (4.2)?

Given $\rho,k\in\mathbb{R}$, define the *generalized tail kernel* associated with a distribution G on $[-\infty,\infty]$ as the transition function $\kappa_G:(0,\infty)\times\mathcal{B}[-\infty,\infty]\to[0,1]$ given by

$$\kappa_G(y\,,\,A) = G\big(y^{-\rho}[A-\psi(y)]\big), \tag{4.3}$$

where ψ is specified in (2.2) (p. 2). Note that κ_G describes transitions between two different spaces. Since ψ satisfies $\psi(uy)=u^\rho\psi(y)+\psi(u)$, a kernel κ has the form (4.3) iff

$$\kappa(uy\,,\,A) = \kappa\big(y\,,\,u^{-\rho}[A-\psi(u)]\big).$$

Proposition 4.1. *Let $K:(0,\infty)\times\mathcal{B}[-\infty,\infty]\to[0,1]$ be a transition function satisfying* (4.1)*, where G is non-degenerate. There exists a family of non-degenerate distributions $\{G_u: 0<u<\infty\}$ on $[-\infty,\infty]$ such that*

$$K\big(tu\,,\,[-\infty,\alpha(t)x+\beta(t)]\big) \Rightarrow G_u([-\infty,x]) \quad \text{on } [-\infty,\infty],\quad 0<u<\infty, \tag{4.4}$$

as $t\to\infty$ if and only if $\alpha,\beta\in\mathrm{ERV}_{\rho,k}$ as in (2.1) *(p. 2). In this case, $G_1=G$, and*

$$K\big(tu_t\,,\,[-\infty,\alpha(t)x+\beta(t)]\big) \Rightarrow \kappa_G(u\,,\,[-\infty,x]) \quad \text{on } [-\infty,\infty] \tag{4.5}$$

whenever $u_t=u(t)\to u\in(0,\infty)$, i.e., the limit is a transition function of the form (4.3)*, where ρ,k are the ERV parameters of α,β.*

Proof. Assume first that $\alpha,\beta\in\mathrm{ERV}_{\rho,k}$ and define

$$h_t(y;u) = \frac{\alpha(tu)}{\alpha(t)}y + \frac{\beta(tu)-\beta(t)}{\alpha(t)},$$

so that by (2.1), $h_t(y_t;u)\to h(y;u)=u^\rho y+\psi(u)$ whenever $y_t\to y\in\mathbb{R}$. For $u>0$,

$$K\big(tu\,,\,[-\infty,\alpha(t)x+\beta(t)]\big) = K\big(tu\,,\,\alpha(tu)\{h_t^{-1}(\cdot\,;u)[-\infty,x]\}+\beta(tu)\big),$$

and applying the second continuous mapping theorem [1] to the weak convergence (4.1), we have

$$K\big(tu\,,\,[-\infty,\alpha(t)x+\beta(t)]\big) \Rightarrow (G\circ h^{-1}(\cdot\,;u))([-\infty,x]).$$

Hence, (4.4) holds with $G_u=\kappa_G(u,\cdot)$, and so $G_1=G$. Furthermore, we have $h_t(x_t;u_t)\to h(x;u)$ whenever $u_t\to u>0$, establishing (4.5).

For the converse, we employ convergence of types. Denote by H_t the distribution $K(t,\cdot)$. Then, on the one hand, we have $H_t([-\infty,\alpha(t)x+\beta(t)])\Rightarrow G_1([-\infty,x])$. On the other hand, fixing $c>0$, we have

$$H_t(\alpha(tc)x+\beta(tc)) = K\big((tc)c^{-1}\,,\,[-\infty,\,\alpha(tc)x+\beta(tc)]\big) \Rightarrow G_{c^{-1}}([-\infty,x]).$$

Convergence of types yields that $\alpha, \beta \in \mathrm{ERV}_{\rho,k}$, and

$$G_{c^{-1}}([-\infty, x]) = G_1([-\infty, c^\rho x + \psi(c)]),$$

with ψ as in (2.2). Using the identity (2.3) (p. 2), we find that G_u has the form (4.3), with $G = G_1$. □

A consequence of Proposition 4.1 is that, in order to obtain a CEVM using (4.1), a necessary and sufficient condition is that α, β are ERV. Requiring G to be non-degenerate is necessary in order to apply convergence of types. For now, we continue to assume that Y is in the standardized domain of attraction.

Theorem 4.1. *Suppose (X, Y) is a random vector on $\mathbb{R} \times [0, \infty)$ and (3.1) holds so that $F_Y \in D(G_1^*)$. Assume $K(y, \cdot) = \mathsf{P}[X \in \cdot \mid Y = y]$ converges according to (4.1) for scaling and centering functions $\alpha(t) > 0$ and $\beta(t) \in \mathbb{R}$ and non-degenerate limit distribution G on $[-\infty, \infty)$. As $t \to \infty$,*

$$t\,\mathsf{P}\left[\left(\frac{X - \beta(t)}{\alpha(t)}, \frac{Y}{t}\right) \in \cdot\right] \xrightarrow{v} \mu(\cdot) \not\equiv 0 \qquad \text{in } \mathbb{M}_+([-\infty, \infty] \times (0, \infty]) \tag{4.6}$$

where μ satisfies (2.9), if and only if $\alpha, \beta \in \mathrm{ERV}_{\rho,k}$. In this case, μ is specified by

$$\mu\left([-\infty, x] \times (y, \infty]\right) = \int_{(y,\infty]} \nu_1(du)\,\mathsf{P}\left[\xi \le u^{-\rho}(x - \psi(u))\right], \quad x \in \mathbb{R},\ y > 0, \tag{4.7}$$

with ψ as in (2.2) (p. 2) and $\xi \sim G$. The expression (4.7) is continuous in x and y if $(\rho, k) \ne (0, 0)$, or if G is continuous.

Proof. The convergence (4.6) to a limit μ satisfying (2.9) implies $\alpha, \beta \in \mathrm{ERV}$ [9, Proposition 1]. Conversely, if $\alpha, \beta \in \mathrm{ERV}_{\rho,k}$, then the convergence (4.6) follows from Lemma **??** (p. **??**) in light of (4.5), yielding the limit in (4.7). We check that $\mu([-\infty, x] \times (y, \infty])$ is continuous when $(\rho, k) \ne (0, 0)$. Indeed, applying dominated convergence, if $x_n \to x$, then

$$\mathsf{P}[\xi \le u^{-\rho}(x_n - \psi(u))] \to \mathsf{P}[\xi \le u^{-\rho}(x - \psi(u))]$$

for all except a countable number of u corresponding to discontinuities of the distribution function. Continuity in y is clear. Also, if $(\rho, k) = (0, 0)$, then $\mu([-\infty, x] \times (y, \infty]) = y^{-1} G([-\infty, x])$, which is continuous if G is. In either case, $\mu([-\infty, x] \times (y, \infty])$ is non-degenerate in x because G is non-degenerate. Finally, $\mu(\{\infty\} \times (y, \infty]) = y^{-1} G(\{\infty\}) = 0$. Therefore, μ satisfies (2.9). □

The limit measure is

$$\mu([-\infty, x] \times (y, \infty]) = \int_0^{y^{-1}} du\,\mathsf{P}\left[\xi \le u^\rho x + \psi(u)\right], \tag{4.8}$$

where

$$u^\rho x + \psi(u) = \begin{cases} u^\rho(x + k\rho^{-1}) - k\rho^{-1} & \rho \ne 0 \\ x + k \log u & \rho = 0 \end{cases}.$$

Changing variables, we obtain the following expressions for μ according to (ρ, k):

$$\mu([-\infty, x] \times (y, \infty]) = \tag{4.9}$$

$$\begin{cases} \dfrac{1}{\rho |x + k\rho^{-1}|^{1/\rho}} \displaystyle\int_0^{|x + k\rho^{-1}| y^{-\rho}} u^{(1-\rho)/\rho}\,\mathsf{P}\left[\xi \le u \operatorname{sgn}(x + k\rho^{-1}) - k\rho^{-1}\right] du & \rho \ne 0 \\ \dfrac{1}{|k| e^{x/k}} \displaystyle\int_{-\infty}^{x \operatorname{sgn}(k) - |k| \log y} e^{u/|k|}\,\mathsf{P}\left[\xi \le u \operatorname{sgn}(k)\right] du & \rho = 0,\quad k \ne 0 \\ y^{-1}\,\mathsf{P}[\xi \le x] & \rho = 0,\quad k = 0 \end{cases}.$$

Here $\operatorname{sgn}(v) = v/|v|\,\mathbf{1}_{\{v\neq 0\}}$, and we read the measure as $y^{-1}\,\mathsf{P}[\xi \leq -k\rho^{-1}]$ when $x = -k\rho^{-1}$ for the case $\rho \neq 0$. Continuity in x and y when $(\rho, k) \neq (0,0)$ is apparent from the above expressions.

We now demonstrate a case where K satisfies (4.1), but (4.6) fails because α, β are not ERV.

Example 4.1. Consider $Y \sim \text{Pareto}(1)$, and $U \sim \text{Uniform}(0,1)$, independent of Y. Put $X = Ue^Y$. Then

$$K(y\,,\,[0,x]) = \mathsf{P}[X \leq x \,|\, Y = y] = \mathsf{P}[U \leq xe^{-y}] = xe^{-y} \wedge 1.$$

In this case, polynomial scaling is not strong enough to give an informative limit, since

$$K(t\,,\,t^\rho[0,x]) = x^\rho t^\rho e^{-t} \wedge 1 \to 0.$$

The appropriate normalization would be exponential $\alpha(t) = e^t$:

$$K(t\,,\,\alpha(t)[0,x]) = xe^t e^{-t} \wedge 1 \to x \wedge 1 = G([0,x]).$$

In fact, by the convergence to types theorem, this the only normalization yielding a non-degenerate limit, up to asymptotic equivalence. However, since α is not regularly varying, Theorem 4.1 shows that (X,Y) cannot follow a CEVM. Indeed, consider

$$\begin{aligned} t\,\mathsf{P}[X \leq \alpha(t)x, Y > ty] &= \int_{(y,\infty]} \nu_1(du) K(tu\,,\,[0, e^t x]) \\ &= \int_{((y\vee 1),\infty]} \nu_1(du)\{xe^{-t(u-1)} \wedge 1\} + \mathbf{1}_{\{y<1\}} \int_{(y,1]} \nu_1(du)\{xe^{-t(u-1)} \wedge 1\}. \end{aligned}$$

The first integral in the previous sum is bounded by $xy^{-1}e^{-t(y-1)} \to 0$. If $y \leq 1$, the second integral approaches $\nu_1(y,1] = y^{-1} - 1$. Therefore, the limit is degenerate in x, violating conditional non-degeneracy (2.9).

In fact, we find that no choice of ERV normalization will lead to a CEVM. Indeed, suppose $\tilde{\alpha}, \tilde{\beta}$ are ERV. Then

$$\begin{aligned} t\,\mathsf{P}[X \leq \tilde{\alpha}(t)x + \tilde{\beta}(t), Y > ty] &= \int_{(y,\infty]} \nu_1(du) K\big(tu\,,\,[0, \tilde{\alpha}(t)x + \tilde{\beta}(t)]\big) \\ &= \int_{(y,\infty]} \nu_1(du)\big\{e^{-tu}(\tilde{\alpha}(t)x + \tilde{\beta}(t)) \wedge 1\big\} \longrightarrow 0 \end{aligned}$$

(see Section 2.1 (p. 2) for a summary of the asymptotic properties of ERV functions).

4.2. **Standardization of X.** Das and Resnick [4, Section 3.2] show that in certain cases, it is possible to standardize the X variable. Denote by x^* and x_* the upper and lower endpoints of the distribution of X respectively, i.e.,

$$x^* = \sup\{x : F_X(x) < 1\} \qquad \text{and} \qquad x_* = \inf\{x : F_X(x) > 0\}.$$

We will call $f : (0,\infty) \to (x_*, x^*)$ a *standardization function* if f is monotone and $\lim_{x\to\infty} f(x) \in \{x_*, x^*\}$. Following [4, Section 3], we will restrict attention to standardization by such functions. However, we have inverted the definition given in [4], in the sense that we will be using $f^\leftarrow$ to standardize rather than f.

For the purpose of this section, we extend the definition of $f^\leftarrow$ in order to invert right-continuous monotone functions which are either increasing or decreasing. Define

$$f^\leftarrow(x) = \begin{cases} \inf\{y : f(y) \geq x\} & \text{if } f \text{ is non-decreasing} \\ \inf\{y : f(y) \leq x\} & \text{if } f \text{ is non-increasing} \end{cases}.$$

Note that $f^{\leftarrow}$ is left-continuous for f non-decreasing and right-continuous for f non-increasing. The main property we shall be using is that

$$\begin{cases} f^{\leftarrow}(x) \le y \iff x \le f(y) & f \text{ non-decreasing} \\ f^{\leftarrow}(x) \le y \iff x \ge f(y) & f \text{ non-increasing} \end{cases}.$$

The distinction between the two cases is a technicality which should not cause confusion in the following discussion. Also, we will say that a monotone function f has two "points of change" if there exist $x_1 < x_2 < x_3$ such that $f(x_1) < f(x_2) < f(x_3)$ for f non-decreasing, and with the opposite inequalities in the non-increasing case.

If (X, Y) satisfy (4.6) for some $\alpha > 0$ and β, then we say (X, Y) can be *standardized* if there exists a standardization function f such that

$$t\,\mathsf{P}\left[t^{-1}(f^{\leftarrow}(X), Y) \in \cdot\,\right] \xrightarrow{v} \mu^*(\cdot) \qquad \text{in } \mathbb{M}_+([0,\infty] \times (0,\infty]), \tag{4.10}$$

where μ^* is a non-null Radon measure. If the limit μ in (4.6) satisfies the conditional non-degeneracy conditions (2.9), then standardization is possible if and only if $(\rho, k) \ne (0, 0)$, i.e., μ is not a product. Because of the dependence on α and β, we can characterize functions f yielding (4.10) in the following way.

Proposition 4.2. *Suppose (X, Y) follow a CEVM, i.e., (4.6) holds with μ satisfying the conditional non-degeneracy conditions (2.9), and $(\rho, k) \ne (0, 0)$. Then a standardization function f standardizes (X, Y) in the sense of (4.10), where μ^* satisfies the conditional non-degeneracy conditions, if and only if*

$$\frac{f(tx) - \beta(t)}{\alpha(t)} \longrightarrow \varphi(x), \qquad x > 0, \tag{4.11}$$

where φ has at least two points of change. In this case, μ and μ^* are related by

$$\mu^*([0, x] \times (y, \infty]) = \mu(A_\varphi(x) \times (y, \infty]),$$

where

$$A_\varphi(x) = \begin{cases} [-\infty, \varphi(x)] & f \text{ non-decreasing} \\ [\varphi(x), \infty] & f \text{ non-increasing} \end{cases}. \tag{4.12}$$

It follows that $\alpha, f \in \mathrm{ERV}$, although not necessarily with the same parameters as α, β. However, depending on the case, f can be expressed in terms of either β or α (see [3, Proposition 2.3.3]).

Proof. Suppose f is non-decreasing. Then for $x, y > 0$, we can write

$$t\,\mathsf{P}\left[\frac{f^{\leftarrow}(X)}{t} \le x, \frac{Y}{t} > y\right] = t\,\mathsf{P}\left[\frac{X - \beta(t)}{\alpha(t)} \le \frac{f(tx) - \beta(t)}{\alpha(t)}, \frac{Y}{t} > y\right]. \tag{4.13}$$

If f satisfies (4.11), then (4.10) holds with

$$\mu^*([0, x] \times (y, \infty]) = \mu([-\infty, \varphi(x)] \times (y, \infty])$$

non-degenerate in x. On the other hand, if (4.10) holds, then (4.13) implies (4.11), and φ has at least two points of increase because μ^* is non-degenerate in x. The mass at $\{\infty\}$ condition in (2.9) follows from the fact that $\lim_{x\to\infty} \varphi(x) = \infty$ if f is non-decreasing (see (4.14) below). The case for f non-increasing is similar, after reversing the inequality for X on the right-hand side of (4.13). $\square$

Assuming (4.11) and $\alpha, \beta \in \mathrm{ERV}_{k,\rho}$, write

$$\frac{f(tx) - \beta(t)}{\alpha(t)} = \frac{\alpha(tx)}{\alpha(t)} \frac{f(tx) - \beta(tx)}{\alpha(tx)} + \frac{\beta(tx) - \beta(t)}{\alpha(t)}$$

and $c = \varphi(1)$. We find that φ has the form

$$\varphi(x) = \begin{cases} cx^\rho + k\rho^{-1}(x^\rho - 1) & \rho \neq 0 \\ c + k\log x & \rho = 0 \end{cases}. \tag{4.14}$$

That φ is non-constant imposes the constraint that $c \neq 0$ if $\rho \neq 0$, $k = 0$.

What if the conditional distribution of X given Y in fact satisfies the kernel convergence assumption (4.1)? We can then apply any standardization directly the conditional distribution via its transition function.

Proposition 4.3. *Suppose the transition function $K : (0,\infty) \times \mathcal{B}[-\infty,\infty] \to [0,1]$ satisfies (4.1) for a probability distribution G on $[-\infty,\infty)$. If f is a standardization function satisfying (4.11), then the transition function $K_f : (0,\infty) \times \mathcal{B}[0,\infty] \to [0,1]$ defined as*

$$K_f(y\,,\, A) = K(y\,,\, f(A))$$

satisfies

$$K_f(t\,,\, t[0,x]) \Rightarrow G(A_\varphi(x)) =: G_f([0,x]) \qquad \text{on } [0,\infty],$$

with $A_\varphi(x)$ as in (4.12). Conversely, suppose $G([0,\infty)) = 1$, and

$$K(t\,,\, t[0,x]) \Rightarrow G([0,x]) \qquad \text{on } [0,\infty]. \tag{4.15}$$

Then, given ERV functions $\alpha > 0$, $\beta \in \mathbb{R}$ on $(0,\infty)$, if f is a monotone function defined on $(0,\infty)$ satisfying (4.11), the transition function $\overline{K}_f : (0,\infty) \times \mathcal{B}[-\infty,\infty] \to [0,1]$ given by

$$\overline{K}_f(y\,,\, A) = K(y\,,\, f^{\leftarrow}(A))$$

satisfies

$$\overline{K}_f\big(t\,,\, [-\infty, \alpha(t)x + \beta(t)]\big) \Rightarrow G(A_{\varphi^{\leftarrow}}(x)) =: \overline{G}_f([-\infty,x]) \qquad \text{on } f([0,\infty]),$$

where

$$A_{\varphi^{\leftarrow}}(x) = \begin{cases} [0, \varphi^{\leftarrow}(x)] & f \text{ non-decreasing} \\ [\varphi^{\leftarrow}(x), \infty] & f \text{ non-increasing} \end{cases}.$$

Proof. Assume (4.1) first, and let f be a non-decreasing standardization function satisfying (4.11). Then,

$$\begin{aligned} K_f\big(t\,, t[0,x]\big) &= K\big(t\,,\, [-\infty, f(tx)]\big) \\ &= K\left(t\,,\, \alpha(t)\left[-\infty, \frac{f(tx) - \beta(t)}{\alpha(t)}\right] + \beta(t)\right) \Rightarrow G([-\infty, \varphi(x)]). \end{aligned}$$

On the other hand, if f satisfies (4.11) for $\alpha, \beta \in \mathrm{ERV}$, then inverting this relation yields

$$\frac{f^{\leftarrow}(\alpha(t)x + \beta(t))}{t} \longrightarrow \varphi^{\leftarrow}(x), \qquad x \in f((0,\infty)).$$

Consequently, for non-decreasing f,

$$\overline{K}_f\big(t\,,\, [-\infty, \alpha(t)x + \beta(t)]\big) = K\big(t\,,\, t[0, t^{-1}f^{\leftarrow}(\alpha(t)x + \beta(t))]\big) \Rightarrow G([0, \varphi^{\leftarrow}(x)]).$$

The case for non-increasing f is similar. □

Proposition 4.3 rephrases Das's result [3, Proposition 2.3.3] on the existence of standardization functions in terms of conditional distributions rather than joint distributions, Indeed, suppose K is a version of the conditional distribution $\mathsf{P}[X \in \cdot \,|\, Y = y]$, satisfying (4.1), where $\alpha, \beta \in \mathrm{ERV}_{\rho,k}$ with $(\rho, k) \neq (0,0)$. If F_Y is in the standardized domain of attraction, then (X,Y) follows a CEVM by Theorem 4.1. Furthermore, (X,Y) can be standardized in the sense of (4.10) [3, Proposition 2.3.3 (1)], and the standardization function f satisfies (4.11) by Proposition 4.2. This standard CEVM could equally have been obtained by applying Theorem 3.1 to the transition function K_f, a version of the conditional distribution $\mathsf{P}[f^{\leftarrow}(X) \in \cdot \,|\, Y = y]$, giving (4.10) directly. That (X, Y) follow a (non-standard) CEVM would then follow by unstandardizing the limit measure μ^*.

In Section 3.1 we discussed the properties of the limit measure μ in the standard case. In particular, if X belongs to the standardized domain of attraction, then G necessarily satisfies the moment condition $0 \le \mathsf{E}\,\xi \le 1$ (recall $\xi \sim G$). This is due to the fact that $\mu((1,\infty] \times (0,\infty])$ is bounded by x^{-1}. Using the standardization approach discussed above, we can derive necessary conditions on G when X belongs to a general domain of attraction.

Suppose there exist normalizing functions $c(t) > 0$ and $d(t)$ such that

$$t\,\mathsf{P}\left[\frac{X - d(t)}{c(t)} > x\right] \longrightarrow (1 + \lambda x)^{-1/\lambda} \qquad x \in \mathbb{E}_\lambda, \tag{4.16}$$

implying that $c, d \in \mathrm{ERV}_{\lambda,1}$ (see Section 2.2, p. 3). Das and Resnick [4, Proposition 4.1] show that if (X, Y) follow a CEVM and (4.16) holds, then the vector (X, Y) belongs to a multivariate domain of attraction provided $\lim_{t\to\infty} \alpha(t)/c(t) \in [0,\infty)$.

For simplicity, we consider the case where the conditional distribution of X given Y satisfies

$$K\big(t\,,\, [-\infty, c(t)x + d(t)]\big) \Rightarrow G([-\infty, x]),$$

i.e., (4.1) holds under the same normalization as in (4.16). Then d is a standardization function satisfying (4.11), and

$$\varphi(x) = \begin{cases} \lambda^{-1}(x^\lambda - 1) & \lambda \ne 0 \\ \log x & \lambda = 0 \end{cases}$$

(see (2.7), p. 3). Theorem 3.1 gives a standard CEVM for $(d^{\leftarrow}(X), Y)$, and furthermore, $d^{\leftarrow}(X)$ belongs to the standardized domain of attraction. Therefore, the distribution G must satisfy

$$\int_0^\infty \mathsf{P}[\xi > \varphi(x)]dx \le 1.$$

Depending on λ, this reduces to

$$\begin{cases} \mathsf{E}\,\xi^{1/\lambda}\,\mathbf{1}_{\{\xi>0\}} \le \lambda^{-1/\lambda} & \lambda > 0 \\ \mathsf{E}(-1/\xi)^{1/|\lambda|}\,\mathbf{1}_{\{\xi<0\}} \le |\lambda|^{1/|\lambda|} & \lambda < 0\,. \\ \mathsf{E}\,e^{\xi} \le 1 & \lambda = 0 \end{cases}$$

Thus, we obtain a different condition for each class of extreme value distribution. In the Fréchet case, we have a bound on the $1/\lambda$-th moment of the right tail. If the domain of attraction is Weibull, this becomes an integrability condition near 0. Finally, in the Gumbel case, the right tail of ξ is exponentially bounded, so all right-tail moments exist.

4.3. **Relation to the Heffernan and Tawn Model.** The CEVM of Theorem 4.1 is inspired by the statistical model proposed by Heffernan and Tawn [10]. We now discuss some links between this work and the CEVM.

Where Heffernan and Tawn's model is based on the convergence of conditional distributions, as in (4.1), the CEVM focuses on limits of joint distributions. Theorem 4.1 shows that Heffernan and Tawn's assumption [10, Equation (3.1)] leads to a CEVM provided the normalization functions α and β are ERV. The fact that the convergence (4.1) is required to hold at all points x suggests that they are expecting a continuous limit. Instead, we have framed the assumption in the more theoretically appealing context of weak convergence. Also, Heffernan and Tawn standardize the conditioning variable to a Gumbel domain of attraction rather than Fréchet, which is our condition (3.1), but this is a minor point.

Also, Example 3.7 demonstrates a theoretical disadvantage to working with conditional distributions. A condition such as (4.1) is tacitly, if not explicitly, assuming a particular version of the conditional distribution. This issue cannot be ignored, since Example 3.7 shows that (4.1) holding for one particular version does not imply that it holds for every version. This question of version is not addressed by Heffernan and Tawn. However, it should not pose a problem if we assume that the distributions are absolutely continuous, as is common in statistical contexts.

Another interesting point concerns the normalization functions α and β. If a non-degenerate CEVM holds, then these are necessarily ERV. It is not clear whether Heffernan and Tawn recognized this as a theoretical result, but they do assume a parametric form for these functions which is very similar to ERV. They specify

$$\alpha(y) = b_{|i}(y) := y^{b_{|i}} = y^{\rho}$$

for some constant $\rho < 1$ and

$$\beta(y) = a_{|i}(y) := \begin{cases} ay & 0 \le \rho < 1, \quad \text{with } a \in [0,1] \\ c - d\log y & \rho < 0 \quad \text{with } a = 0,\ c \in \mathbb{R},\ d \in [0,1] \end{cases}.$$

Although more general models are possible, the form of the ERV limit function ψ in (2.2) (p. 2) suggests that a parametric approach is indeed reasonable.

5. General Normalizations for both X and Y

Up until now, we have been assuming that Y belongs to the standardized Fréchet domain of attraction: $t\,\mathsf{P}[Y > ty] \to y^{-1}$ for $y > 0$. We wish to extend the result of Theorem 4.1 to the case where Y belongs to a general domain of attraction:

$$t\,\mathsf{P}[Y > a(t)y + b(t)] \longrightarrow (1+\gamma y)^{-1/\gamma} \qquad y \in \mathbb{E}_\gamma, \tag{5.1}$$

where $\mathbb{E}_\gamma := \{y : 1+\gamma y > 0\}$. See Section 2.2 (p. 3) for further details on domains of attraction. We will assume that $b(t)$ is given by (2.6).

An important consideration in the previous development is that convergence results depend on properties of the particular choice of version $K(y,\cdot)$ of the conditional distribution $\mathsf{P}[X \in \cdot \,|\, Y = y]$. Because Y is now normalized according to a and b, the condition (4.1) may no longer be sufficient to obtain a general CEVM limit, as in (2.8) (p. 3).

If it were known that $K\big(a(t)u + b(t)\,,\,[-\infty, \alpha(t)x + \beta(t)]\big) \to \varphi_x(u)$ for $u > 0$, then (2.8) should follow from arguments similar to those in Section 4.1. On the other hand, Heffernan and Resnick [9] argue that (2.8) reduces to (4.6) by standardizing Y using the transformation $Y \mapsto b^{\leftarrow}(Y)$. Hence, if K^*, a specific version of $\mathsf{P}[X \in \cdot \,|\, b^{\leftarrow}(Y) = y]$, satisfies (4.1), then $(X, b^{\leftarrow}(Y))$ follows a CEVM under approriate normalization of X by Theorem 4.1, and (2.8) should follow from (4.6) by untransforming. We now examine the consistency of these two approaches.

5.1. **Kernel Asymptotics.** The transition function $K : (-\infty,\infty) \times \mathcal{B}[-\infty,\infty] \to [0,1]$ will continue to denote a specific version of the conditional distribution of X given Y, i.e., ,

$$K(y\,,\cdot) = \mathsf{P}[X \in \cdot \,|\, Y = y].$$

Moving towards the transformation approach described above, we first argue that we can express a version of the conditional distribution of X given $b^{\leftarrow}(Y)$ in terms of K.

First, recall that the convergence (5.1), where b is given by (2.6), implies that $a, b \in \mathrm{ERV}_{\gamma,1}$. Hence, $a \in \mathrm{RV}_\gamma$, and

$$\frac{b(tx) - b(t)}{a(t)} \longrightarrow \begin{cases} \dfrac{x^\gamma - 1}{\gamma} & \gamma \neq 0 \\ \log x & \gamma = 0 \end{cases}, \qquad x > 0. \tag{5.2}$$

Inverting (5.2) gives

$$\frac{b^{\leftarrow}(a(t)x+b(t))}{t} \longrightarrow \begin{cases}(1+\gamma x)^{1/\gamma} & \gamma \neq 0 \\ e^{x} & \gamma = 0\end{cases}, \qquad x \in \mathbb{E}_{\gamma}. \tag{5.3}$$

Furthermore, recall that if $\tilde{b}$ is any function on $(0,\infty)$ satisfying

$$\frac{\tilde{b}(t)-b(t)}{a(t)} \longrightarrow 0 \qquad \text{as } t \to \infty, \tag{5.4}$$

then (5.1), (5.2), and (5.3) hold with b replaced by $\tilde{b}$. We now verify that we can choose such a $\tilde{b}$ which is invertible.

Lemma 5.1. *There exists a function b^* satisfying* (5.4) *that is continuous and strictly monotone.*

Proof. We consider cases on γ. If $\gamma = 0$, then $b \in \Pi(a)$. Then we can find $\bar{b}$ continuous, strictly increasing such that $(\bar{b}(t) - b(t))/a(t) \to 1$ by [12, Proposition 0.16]. The choice $b^*(x) = \bar{b}(e^{-1}x)$ satisfies (5.4). Otherwise, suppose $\gamma > 0$. Then $b \in \mathrm{RV}_{\gamma}$, and $b(t)/a(t) \to \gamma^{-1}$ [6, Theorem B.2.2 (1)]. Consequently, [13, Proposition 2.6 (vii)] gives a continuous, strictly increasing function $b^* \sim b$. Writing

$$\frac{b^*(t)-b(t)}{a(t)} = \frac{b(t)}{a(t)}\left[\frac{b^*(t)}{b(t)} - 1\right]$$

shows that b^* satisfies (5.4). Finally, if $\gamma < 0$, then $b(\infty) = \lim_{t\to\infty} b(t)$ exists finite, $b(\infty) - b \in \mathrm{RV}_{\gamma}$, and $(b(\infty) - b(t))/a(t) \to -\gamma^{-1}$. Choose $\bar{b}$ continuous, strictly decreasing, with $\bar{b} \sim (b(\infty) - b)$, and set $b^* = b(\infty) - \bar{b}$. $\square$

Henceforth, b^* will denote a continuous, strictly monotone function satisfying (5.4). The advantage to working with b^* is that $b^{*\leftarrow}(b^*(x)) = b^*(b^{*\leftarrow}(x)) = x$. By (5.2), $Y^* = b^{*\leftarrow}(Y)$ belongs to the standard domain of attraction when (5.1) holds:

$$t\,\mathsf{P}[Y^* > ty] = t\,\mathsf{P}\left[\frac{Y - b^*(t)}{a(t)} > \frac{b^*(ty) - b^*(t)}{a(t)}\right] \longrightarrow y^{-1}, \qquad y > 0.$$

We argue that when $K(y,\cdot) = \mathsf{P}[X \in \cdot \mid Y = y]$, the transition function

$$K^*(y\,,\,\cdot) := K(b^*(y)\,,\,\cdot) \tag{5.5}$$

is a version of the conditional distribution $\mathsf{P}[X \in \cdot \mid Y^* = y]$.

Proposition 5.1. *For measurable A and $y > 0$, we have*

$$\mathsf{P}[X \in A, Y^* > y] = \int_{(y,\infty)} K(b^*(u)\,,\,A)\,\mathsf{P}[Y^* \in du]. \tag{5.6}$$

Proof. Write

$$\begin{aligned}\mathsf{P}[X \in A, Y^* > y] = \mathsf{P}[X \in A, Y > b^*(y)] &= \int_{(b^*(y),\infty)} K(u\,,\,A)\,\mathsf{P}[Y \in du] \\ &= \int_{(b^*(y),\infty)} K\big(b^*(b^{*\leftarrow}(u))\,,\,A\big)\,\mathsf{P}[Y \in du],\end{aligned}$$

using the fact that $b^*(b^{*\leftarrow}(u)) = u$ for all u, and change variables according to the transformation $T = b^{*\leftarrow}$. Since $T^{-1}(y,\infty) = \{x : b^{*\leftarrow}(x) > y\} = (b^*(y),\infty)$, the result follows. $\square$

Note that (5.6) is not necessarily true for the function b given by (2.6), unless $\mathsf{P}[b(b^{\leftarrow}(Y)) \neq Y] = 0$.

Next, we show that the two approaches to the CEVM discussed at the beginning of Section 5, the direct approach and the standardization approach, are indeed consistent. That K^* converges to a family of distributions under scaling of the initial state, in the sense of (4.4) (p. 9), is equivalent to the same for K with initial state normalized by a and b.

Proposition 5.2. *Suppose Y is a random variable with distribution satisfying* (5.1)*, and let K^* be given by* (5.5)*. Given normalization functions $\alpha(t) > 0$ and $\beta(t) \in \mathbb{R}$, there exists a transition function $\phi^* : (0,\infty) \times \mathcal{B}[-\infty,\infty] \to [0,1]$ such that, as $t \to \infty$,*

$$K^*\big(tu_t\,,\, [-\infty, \alpha(t)x + \beta(t)]\big) \Rightarrow \phi^*(u\,,\, [-\infty, x]) \qquad \text{on } [-\infty,\infty] \tag{5.7}$$

whenever $u_t \to u \in (0,\infty)$, if and only if there exists a transition function $\phi : \mathbb{E}_\gamma \times \mathcal{B}[-\infty,\infty] \to [0,1]$ such that, as $t \to \infty$,

$$K\big(a(t)u_t + b(t)\,,\, [-\infty, \alpha(t)x + \beta(t)]\big) \Rightarrow \phi(u\,,\, [-\infty, x]) \qquad \text{on } [-\infty,\infty] \tag{5.8}$$

whenever $u_t \to u \in \mathbb{E}_\gamma$. If these convergences hold, then

(i) $\alpha, \beta \in \mathrm{ERV}$*;*
(ii) $\phi^* = \kappa_{G^*}$*, a generalized tail kernel* (4.3) *with* $G^* = \phi^*(1,\cdot)$*;*
(iii) $\phi(u\,,\, A) = \kappa_G((1+\gamma u)^{1/\gamma}\,,\, A)$*, where κ_G is a generalized tail kernel with $G = \phi(0,\cdot)$; and*
(iv) *the two transition functions are related by $G = G^*$.*

Proof. Abbreviate $a_t = a(t)$ and $b_t = b(t)$. The convergences (5.2) and (5.3) are in fact locally uniform on $(0,\infty)$ (see Section 2.1, p. 2). Since b^* satisfies (5.4), it follows that

$$\frac{b^*(tu_t) - b_t}{a_t} \longrightarrow \frac{u^\gamma - 1}{\gamma} \qquad \text{whenever } u_t \to u \in (0,\infty),$$

and

$$\frac{b^{*\leftarrow}(a_t u_t + b_t)}{t} \longrightarrow (1+\gamma u)^{1/\gamma} \qquad \text{whenever } u_t \to u \in \mathbb{E}_\gamma.$$

Assuming (5.7), for $u_t \to u \in \mathbb{E}_\gamma$ we have

$$\begin{aligned} &K\big(a(t)u_t + b(t)\,,\, [-\infty, \alpha(t)x + \beta(t)]\big) \\ &\quad = K\Big(b^*\big(t\{t^{-1}b^{*\leftarrow}(a_t u_t + b_t)\}\big)\,,\, [-\infty, \alpha(t)x + \beta(t)]\Big) \\ &\quad = K^*\big(t\{t^{-1}b^{*\leftarrow}(a_t u_t + b_t)\}\,,\, [-\infty, \alpha(t)x + \beta(t)]\big) \\ &\quad \Rightarrow \phi^*\big((1+\gamma u)^{1/\gamma}\,,\, [-\infty, x]\big) =: \phi(u\,,\, [-\infty, x]) \end{aligned}$$

Conversely, if (5.8) holds, then for $u_t \to u > 0$,

$$\begin{aligned} &K^*\big(tu_t\,,\, [-\infty, \alpha(t)x + \beta(t)]\big) \\ &\quad = K\big(a_t \cdot a_t^{-1}(b^*(tu_t) - b_t) + b_t\,,\, [-\infty, \alpha(t)x + \beta(t)]\big) \\ &\quad \Rightarrow \phi\big(\gamma^{-1}(u^\gamma - 1)\,,\, [-\infty, x]\big) =: \phi^*(u\,,\, [-\infty, x]) \end{aligned}$$

In either case, $G := \phi(0,\cdot) = \phi^*(1,\cdot) =: G^*$. Proposition 4.1 shows that α and β are ERV and $\phi^* = \kappa_{G^*}$. Consequently, $\phi(u\,,\,\cdot) = \kappa_G((1+\gamma u)^{1/\gamma}\,,\,\cdot)$. □

Therefore, by Proposition 4.1 (p. 9), if there exists a non-degenerate distribution G on $[-\infty,\infty)$ such that

$$K^*\big(t\,,\, [-\infty, \alpha(t)x + \beta(t)]\big) = K\big(b^*(t)\,,\, [-\infty, \alpha(t)x + \beta(t)]\big) \Rightarrow G([-\infty, x]) \tag{5.9}$$

with $\alpha, \beta \in \mathrm{ERV}$, then (5.8) holds.

How can we apply Proposition 5.2 starting from an assumption like (5.9) on the kernel K rather than K^*? Because $b^*(b^{*\leftarrow}(t)) = t$, (5.9) can be written as

$$K\big(t\,,\,[-\infty, \alpha \circ b^{*\leftarrow}(t)x + \beta \circ b^{*\leftarrow}(t)]\big) \Rightarrow G([-\infty, x]) \qquad \text{as } t \to y^*,$$

where y^* denotes the upper endpoint of the distribution of Y, written as $y^* = \sup\{y : F_Y(y) < 1\}$. Therefore, we require there to exist a non-degenerate distribution G and normalization functions $\tilde{\alpha} > 0$ and $\tilde{\beta}$ such that

$$K\big(t\,,\,[-\infty, \tilde{\alpha}(t)x + \tilde{\beta}(t)]\big) \Rightarrow G([-\infty, x]) \qquad \text{as } t \to y^*, \tag{5.10}$$

and $\alpha = \tilde{\alpha} \circ b^*$, $\beta = \tilde{\beta} \circ b^* \in \mathrm{ERV}$.

5.2. **CEVM Properties.** Using the standardization approach discussed in the previous section, we obtain a CEVM when Y belongs to a general domain of attraction.

Theorem 5.1. *Suppose (X, Y) is a random vector on $\mathbb{R}^2$, where $F_Y \in D(G_\gamma)$ (5.1), and $K(y, \cdot) = \mathsf{P}[X \in \cdot \,|\, Y = y]$ converges according to (5.10), for some normalizing functions $\tilde{\alpha} > 0$ and $\tilde{\beta} \in \mathbb{R}$ and non-degenerate limit distribution G on $[-\infty, \infty)$. Let b^* be the function satisfying (5.4) given by Lemma 5.1, and put $\alpha = \tilde{\alpha} \circ b^*$, $\beta = \tilde{\beta} \circ b^*$. Then, as $t \to \infty$,*

$$t\,\mathsf{P}\left[\left(\frac{X - \beta(t)}{\alpha(t)}, \frac{Y - b(t)}{a(t)}\right) \in \cdot\,\right] \xrightarrow{v} \mu(\cdot) \qquad \text{in } \mathbb{M}_+([-\infty, \infty] \times \overline{\mathbb{E}}_\gamma), \tag{5.11}$$

where μ is a non-null Radon measure satisfying the conditional non-degeneracy conditions (2.9), if and only if $\alpha, \beta \in \mathrm{ERV}_{\rho,k}$. In this case, the limit measure μ is specified by

$$\mu([-\infty, x] \times (y, \infty]) = \int_0^{(1+\gamma y)^{-1/\gamma}} du\,\mathsf{P}\left[\xi \le u^\rho x + \psi(u)\right], \qquad x \in \mathbb{R},\ y \in \mathbb{E}_\gamma, \tag{5.12}$$

with ψ as in (2.2) (p. 2). The expression (5.12) is continuous in x and y if $(\rho, k) \neq (0, 0)$.

Proof. First, observe that $Y^* = b^{*\leftarrow}(Y) \in D(G_1^*)$. Defining the transition function $K^*(y, \cdot) = \mathsf{P}[X \in \cdot \,|\, Y^* = y]$ as in (5.5), our hypotheses imply (5.9). Therefore, if $\alpha, \beta \in \mathrm{ERV}_{\rho,k}$, then by Theorem 4.1, we have

$$t\,\mathsf{P}\left[\left(\frac{X - \beta(t)}{\alpha(t)}, \frac{Y^*}{t}\right) \in \cdot\,\right] \xrightarrow{v} \mu^*(\cdot) \qquad \text{in } \mathbb{M}_+([-\infty, \infty] \times (0, \infty]),$$

where μ^* is defined by

$$\mu^*([-\infty, x] \times (y, \infty]) = \int_0^{y^{-1}} du\,\mathsf{P}\left[\xi \le u^\rho x + \psi(u)\right], \quad x \in \mathbb{R},\ y > 0,$$

conditionally non-degenerate. Consequently, for $x \in \mathbb{R}$ and $y \in \mathbb{E}_\gamma$,

$$\begin{aligned}
t\,\mathsf{P}&\left[\frac{X - \beta(t)}{\alpha(t)} \le x, \frac{Y - b(t)}{a(t)} > y\right] \\
&= t\,\mathsf{P}\left[\frac{X - \beta(t)}{\alpha(t)} \le x, \frac{Y^*}{t} > \frac{b^{*\leftarrow}(a(t)y + b(t))}{t}\right] \\
&= \int_0^{(1+\gamma y)^{-1/\gamma}} du\,\mathsf{P}\left[\xi \le u^\rho x + \psi(u)\right] = \mu([-\infty, x] \times (y, \infty]),
\end{aligned}$$

and the marginal transformation of Y does not affect conditional non-degeneracy or continuity. Conversely, (5.11) implies that $\alpha, \beta \in \mathrm{ERV}$ [9, Proposition 1]. □

Remark. Instead of standardizing Y, we could equally have used the convergence (5.8), which holds under our assumptions by Propositions 4.1 and 5.2.

Recalling the forms of the limit measure given in Section 4.1, we can express the limit measure in (5.12) as

$$\mu([-\infty,x]\times(y,\infty]) =$$
$$\begin{cases} \dfrac{1}{\rho|x+k\rho^{-1}|^{1/\rho}}\displaystyle\int_0^{|x+k\rho^{-1}|(1+\gamma y)^{-\rho/\gamma}} u^{(1-\rho)/\rho}\,\mathsf{P}\left[\xi\le u\,\mathrm{sgn}(x+k\rho^{-1})-k\rho^{-1}\right]du & \rho\neq 0\\ \dfrac{1}{|k|e^{x/k}}\displaystyle\int_{-\infty}^{x\,\mathrm{sgn}(k)-|k|\gamma^{-1}\log(1+\gamma y)} e^{u/|k|}\,\mathsf{P}\left[\xi\le u\,\mathrm{sgn}(k)\right]du & \rho=0,\ k\neq 0\\ (1+\gamma y)^{-1/\gamma}\,\mathsf{P}[\xi\le x] & \rho=0,\ k=0\end{cases}$$

where $\mathrm{sgn}(v)=v/|v|\,\mathbf{1}_{\{v\neq 0\}}$, and we read the measure as $(1+\gamma y)^{-1/\gamma}\,\mathsf{P}[\xi\le -k\rho^{-1}]$ when $x=-k\rho^{-1}$ for the case $\rho\neq 0$.

In Example 4.1 (p. 11), we presented a transition function satisfying (4.1) which did not lead to a CEVM when paired with $Y\in D(G_1^*)$. We now show that a non-degenerate CEVM may be obtained if Y belongs to a non-standardized domain of attraction.

Example 5.1. Consider $Y\sim\mathrm{Exp}(1)$, and $U\sim\mathrm{Uniform}(0,1)$, independent of Y. Put $X=Ue^Y$. Note that $Y\in D(G_0)$ with $a(t)\equiv 1$, $b(t)=\log t$, since for $y\in\mathbb{R}$,

$$t\,\mathsf{P}(Y>y+\log t)=te^{-y-\log t}=e^{-y}.$$

A version of the conditional distribution is given by

$$K(y\,,\,[0,x])=\mathsf{P}[X\le x\,|\,Y=y]=\mathsf{P}[U\le xe^{-y}]=xe^{-y}\wedge 1.$$

Taking $\tilde\alpha(t)=e^t$, we saw in Example 4.1 that

$$K(t\,,\,\tilde\alpha(t)[0,x])\Rightarrow x\wedge 1=G([0,x]),$$

although $\tilde\alpha$ is not regularly varying. Since b is continuous and strictly monotone, set $\alpha(t)=\tilde\alpha(b(t))=t$. Then

$$K^*(t\,,\,t[0,x])=K(b(t)\,,\,\tilde\alpha(b(t))[0,x])\Rightarrow G([0,x]),$$

and $\alpha(t)\in\mathrm{RV}_1$. Hence, $K^*(tu\,,\,\alpha(t)[0,x])\Rightarrow xu^{-1}\wedge 1=G(u^{-1}[0,x])$, and $K^*(y,\cdot)=\mathsf{P}[X\in\cdot\,|\,e^Y=y]$. On the other hand, note that for $u\in\mathbb{R}$,

$$K\big(a(t)u+b(t)\,,\,\alpha(t)[0,x]\big)=txe^{-u-\log t}\wedge 1=xe^{-u}\wedge 1=G\big((e^u)^{-1}[0,x]\big).$$

This illustrates the equivalence presented in Proposition 5.2 (p. 17). Now, for $x>0$, $y>0$, the joint distribution is given by

$$\mathsf{P}[X\le x,Y>y]=\int_{\log x\vee y}^{\infty}xe^{-2u}du+\int_y^{\log x}e^{-u}du\,\mathbf{1}_{\{y<\log x\}}\,.$$

Therefore, for $x>0$, $y\in\mathbb{R}$, and large t, we have

$$t\,\mathsf{P}[X\le tx,Y>y+\log t]=\left\{\begin{array}{ll}\dfrac{xe^{-2y}}{2} & \text{if } \log x\le y\\ e^{-y}-\dfrac{1}{2x} & \text{if } \log x>y\end{array}\right\}=\mu([0,x]\times(y,\infty]),$$

and (X,Y) follow a CEVM by Theorem 5.1.

6. Conclusions and Future Directions

Although dealing with conditional distributions in general raises certain issues surrounding identifiability, in many statistical contexts a conditional formulation such as (4.1) is convenient. For example, it may be appropriate to model X as an explicit function of Y. Also, if we are working with distributions that have continuous densities, the natural choice of version of the conditional distribution is the absolutely continuous one, and other simplifications may be afforded.

The above development suggests that in such cases, the approach of Heffernan and Tawn [10] is reasonable, and will generally lead to a fairly parsimonious extremal model which can account for varying degrees of asymptotic independence. Heffernan and Tawn propose a semiparametric model, where the limit distribution G is estimated nonparametrically, and the normalization functions α and β belong to a parametric family. The extended regular variation of α and β provides some justification for this last assumption. Furthermore, the formulas for the limit measure derived above show that by modeling conditional distributions, we obtain a simpler CEV model parametrized by the distribution G and the pair (ρ, k), along with γ, the extreme value index of Y.

The question of fitting a bivariate CEV model has been considered by Das and Resnick [5] and by Fougères and Soulier [8]. These authors discuss statistics for detecting a CEV model and estimating the normalizing functions. However, many open questions remain, such as the asymptotic distributions of such estimators, and the appropriate method for nonparametric estimation of G. These problems may presumably be simplified substantially through the use of standardization for both X and Y.

Also, a natural extension of the bivariate model discussed above would be to consider the conditional formulation for higher-dimensional vectors. Indeed, this was the original intention of Heffernan and Tawn, who apply their methodology to a five-dimensional air pollution dataset. It is not clear what would be the appropriate formulation of a model conditioning on more than one extreme variable, nor the connections between such a model and the usual multivariate domain of attraction. In particular, cases where asymptotic independence is present between some pairs of variables but not others would require careful treatment. However, a model based on conditional distributions as developed above should prove useful.

References

[1] P. Billingsley. *Convergence of Probability Measures.* Wiley, 1968.

[2] N.H. Bingham, C.M. Goldie, and J.L. Teugels. *Regular Variation.* Encyclopedia of Mathematics and Its Applications. Cambridge University Press, 1989.

[3] B. Das. *The conditional extreme value model and related topics.* PhD thesis, Cornell University, 2009.

[4] B. Das and S.I. Resnick. Conditioning on an extreme component: Model consistency with regular variation on cones. *Bernoulli*, 17(1):226–252, 2011.

[5] B. Das and S.I. Resnick. Detecting a conditional extreme value model. *Extremes*, 14(1):29–61, 2011.

[6] L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction.* Springer Verlag, 2006.

[7] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 2. Wiley, New York, 2nd edition, 1971.

[8] A.L. Fougères and P. Soulier. Estimation of conditional laws given an extreme component. *Extremes*, 15(1):1–34, 2012.

[9] J.E. Heffernan and S.I. Resnick. Limit laws for random vectors with an extreme component. *Annals of Applied Probability*, 17(2):537–571, 2007.

[10] J.E. Heffernan and J.A. Tawn. A conditional approach for multivariate extreme values. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 66(3):497–546, 2004.

[11] S.I. Resnick. *A Probability Path.* Birkhäuser, Boston, 1999.
[12] S.I. Resnick. *Extreme Values, Regular Variation, and Point Processes.* Springer Verlag, 2007.
[13] S.I. Resnick. *Heavy-tail Phenomena: Probabilistic and Statistical Modeling.* Springer Verlag, 2007.
[14] S.I. Resnick. Multivariate regular variation on cones: application to extreme values, hidden regular variation and conditioned limit laws. *Stochastics: An International Journal of Probability and Stochastics Processes*, 80(2-3):269–298, 2008.
[15] S.I. Resnick and D. Zeber. Asymptotics of Markov kernels and the tail chain. *Advances in Applied Probability*, 45(1), 2013. To appear.
[16] J. Segers. Multivariate regular variation of heavy-tailed Markov chains. *Arxiv preprint math/0701411*, 2007.
[17] E. Seneta, A. Dold, and B. Eckmann. *Regularly varying functions.* Springer-Verlag Berlin-Heidelberg, 1976.

Sidney I. Resnick, School of ORIE, Rhodes Hall 284, Cornell University, Ithaca, NY 14853
E-mail address: `sir1@cornell.edu`

David Zeber, Department of Statistical Science, 301 Malott Hall, Cornell University, Ithaca NY 14853
E-mail address: `dsz5@cornell.edu`